Metabarcoding and Metagenomics 6: 349–365
DOI 10.3897/mbmg.6.85844

Research Article

Recommendations for the preservation of environmental samples in diatom metabarcoding studies

Ana Baricevic[1], **Cécile Chardon**[2], **Maria Kahlert**[3], **Satu Maaria Karjalainen**[4], **Daniela Maric Pfannkuchen**[1], **Martin Pfannkuchen**[1], **Frédéric Rimet**[2], **Mirta Smodlaka Tankovic**[1], **Rosa Trobajo**[5], **Valentin Vasselon**[2,6], **Jonas Zimmermann**[7], **Agnès Bouchez**[2]

1 *Ruder Boskovic Institute, Center for Marine Research, Rovinj, Croatia*
2 *UMR CARRTEL, INRAE, USMB, Thonon-les-Bains, France*
3 *Swedish University of Agricultural Sciences, Department of Aquatic Sciences and Assessment, Uppsala, Sweden*
4 *Finnish Environment Institute, Oulu, Finland*
5 *IRTA-Institute for Food and Agricultural Research and Technology, Marine and Continental Waters Programme, La Ràpita, Catalonia, Spain*
6 *SCIMABIO-Interface, Thonon-les-Bains, France*
7 *Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Berlin, Germany*

Corresponding author: Agnès Bouchez (agnes.bouchez@inrae.fr)

Academic editor: Kristy Deiner | **Received** 26 April 2022 | **Accepted** 28 September 2022 | **Published** 24 November 2022

Abstract

Implementation of DNA metabarcoding for diatoms for environmental monitoring is now moving from a research to an operational phase, requiring rigorous guidelines and standards. In particular, the first steps of the diatom metabarcoding process, which consist of sampling and storage, have been addressed in various ways in scientific and pilot studies and now need to be rationalised. The objective of this study was to compare three currently applied preservation protocols through different storage durations (ranging from one day to one year) for phytobenthos and phytoplankton samples intended for diatom DNA metabarcoding analysis. The experimental design used samples from four freshwater and two marine sites of diverse ecological characteristics. The impact of the sample preservation and storage duration was assessed through diatom metabarcoding endpoints: DNA quality and quantity, diversity and richness, diatom assemblage composition and ecological index values (for freshwater samples). The yield and quality of extracted DNA only decreased for freshwater phytobenthos samples preserved with ethanol. Diatom diversity was not affected and their taxonomic composition predominantly reflected the site origin. Only rare taxa (< 100 reads) differed among preservation methods and storage durations. For biomonitoring purposes, freshwater ecological index values were not affected by the preservation method and storage duration tested (including ethanol preservation), all treatments returning the same ecological status for a site. This study contributes to consolidating diatom metabarcoding. Thus, accompanied by operational standards, the method will be ready to be confidently deployed and prescribed in future regulatory monitoring.

Key Words

biomonitoring, diatom assemblages, DNA metabarcoding, European Water Framework Directive, methods, sample preservation

Introduction

Aquatic ecosystems provide many ecosystem services and functions, such as fishing, water provisioning and recreation, and are hosts to considerable biodiversity (Grizzetti et al. 2016). However, these ecosystems are subjected to many pressures that can cause physical alteration, water pollution and invasion by non-native species. Faced with these pressures, government agencies have implemented management programs for their ecosystems. In Europe, the Water Framework Directive (WFD, European Commission 2000) and the Marine Strategy Framework Directive (MSFD, European Commission 2008) have been implemented for aquatic ecosystem monitoring.

Copyright *Ana Baricevic et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

These directives set criteria, based on biological communities, to assess the ecological quality of aquatic ecosystems. Diatoms, an abundant group of microalgae, are used as bioindicators in both marine and freshwater ecosystems. In marine ecosystems, plankton diatom assemblages are present in pelagic areas where some toxic species can bloom. Monitoring of diatom assemblages enables the dynamics of such species to be followed. In freshwater ecosystems, such as rivers and lakes, benthic diatoms are used to assess the ecological quality of ecosystems. Diatoms encompass huge taxonomic diversity (Mann and Vanormelingen 2013) and each species has unique ecological preferences making them excellent ecological indicators (Rimet 2011). Different biotic indices based on the ecological preferences of the most abundant species have been developed and standardised for WFD applications in European countries (Kelly 2013).

Until now, standard methods used to count and identify diatoms to species level are based on morphological criteria visible by light microscopy (CEN 2014a,b). This is time-consuming, requires highly-trained taxonomists and can present considerable inter-operator variation (Kahlert et al. 2012). However, during the past decade, the development of DNA metabarcoding coupled with High-Throughput Sequencing (HTS) has offered an alternative (Kermarrec et al. 2013) that can be applied to biomonitoring (e.g. Trobajo et al. 2021). Several studies conducted at regional scales (e.g. two cantons in Switzerland with 87 samples: Apothéloz-Perret-Gentil et al. 2017; Mayotte Island, France, with 80 samples: Vasselon et al. 2017a; Catalonia, Spain, with 160 samples: Pérez-Burillo et al. 2021), at national scale (e.g. France with 450 samples: Rivera et al. 2020) and at transnational scale (e.g. in Fennoscandia: Bailet et al. 2019; along the Danube River with JDS4: Zimmermann et al. 2021) confirmed that this approach is applicable to freshwater biomonitoring purposes. This approach was also recently applied to coastal waters (Pérez-Burillo et al. 2022). Marine plankton diversity has been characterised using metabarcoding within important European initiatives like Tara Oceans (de Vargas et al. 2015) and Biomarks (Massana et al. 2015).

The diatom metabarcoding process involves five steps: 1) sampling and storage, 2) DNA extraction, 3) PCR amplification, 4) amplicon library preparation and sequencing and 5) bioinformatics treatment. All these steps can show variations among studies with a large range of protocols that are used. To date, only a few studies have compared different protocols; Vasselon et al. (2017b) compared different DNA extraction protocols; Kermarrec et al. (2013, 2014) compared different barcodes; Vasselon et al. (2018) applied a cell biovolume correction factor to make quantification by microscopy and metabarcoding comparable; and Tapolczai et al. (2019), Rivera et al. (2020), Bailet et al. (2020) and Kang et al. (2021) compared the impact of different bioinformatics pipelines on biotic indices. None of these studies, however, dealt with the first steps of the metabarcoding process, i.e. sampling and storage.

Standardisation efforts at CEN (European Committee for Standardisation) have accompanied the application of the European Directives although standardisation of genomic methods for biomonitoring is still in its infancy. In 2018, CEN published two technical reports dealing with the management of diatom DNA barcodes (CEN 2018a) and updates to the sampling protocol to enable DNA extraction from the samples (CEN 2018b). However, the preservation methods currently used for morphological analyses aimed at preserving silica frustules and are based on Lugol's iodine or formaldehyde, which do not preserve DNA adequately for subsequent DNA-based applications. Alternative preservation methods have been used in various scientific metabarcoding studies in order to store samples without jeopardising the quality or quantity of DNA that could be extracted from them: deep-freezing (Visco et al. 2015), ethanol (final concentration 70%: Vasselon et al. 2017a; Rimet et al. 2018) and commercial or home-made nucleic acid preservative solutions that rapidly permeate tissues to stabilise and protect cellular RNA/DNA (Kelly et al. 2018, 2020). Based on all these studies, the CEN Technical Report published in 2018 (CEN 2018b) presented this variety of preservation methods as DNA-friendly alternatives. However, little was known about the relative effectiveness of these various approaches to preserve raw samples sustainably. Moreover, to our knowledge, no guidance is available on how long the collection of preserved samples may be stored to ensure reliable results using DNA metabarcoding.

Genomic methods for environmental monitoring are moving from research to operational applications. The choice of a preservation method by the end-users depends on sampling and shipment operational constraints. For example, during a field sampling day, including the visit to several potentially remote sites, deep-freezing may be difficult. If a sample shipment is required, it is cheaper and safer to use a preservative that is free of hazardous compounds (e.g. formaldehyde). Moreover, while several hundred samples can be processed in a single sequencing run, the time to collect all these samples in the field can last weeks to months. So, in order to derive best-practices for developing standards, it is important to know if the preservation protocol and/or the storage duration have an impact on the final assessment of the diatom assemblage.

The aim of our study is to highlight best practices for preserving phytobenthos and phytoplankton samples for DNA-based applications involving diatoms. With that aim, we compared different preservation methods and storage durations. The recommendations obtained will be useful in the context of subsequent standardisation.

We compared three preservation protocols through different storage durations. These are based on those proposed in CEN (2018b) and are described in literature: 1) ethanol, 2) deep freezing or 3) nucleic acid preservative. The effect of storage duration was evaluated through six different storage durations, ranging from one day to one year. Methods were tested on phytoplankton samples from two marine sites and on phytobenthos samples from four contrasting river sites in Europe. Preservation methods were compared over time based on: 1) the quantity and quality of extracted DNA, 2) the diatom assemblage diversity and structure assessed

through DNA metabarcoding (Vasselon et al. 2017a) and 3) and ecological quality indices (freshwater sites).

Materials and methods

The experimental design is summarised in Fig. 1 and described in detail in Suppl. material 1: Data 1.

Site selection and sampling

Six contrasting European sites (two Mediterranean marine sites – Spain, Croatia; four European river sites – France, Spain, Germany, Finland) were selected for sampling, based on differences in water quality and typology (Table 1). Sampling at all sites was done on the same day (18 September 2017).

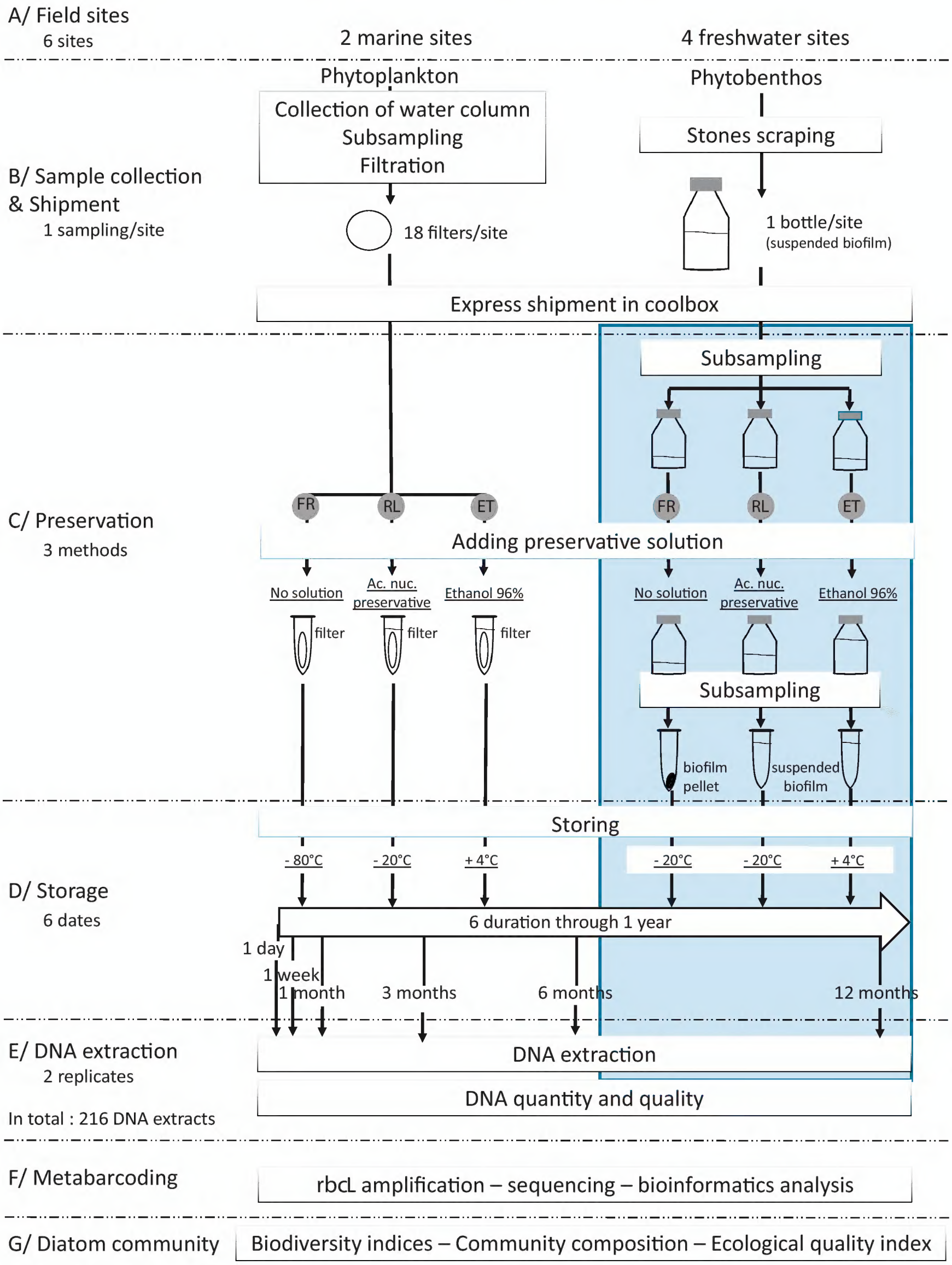

Figure 1. Workflow of the study presenting the three preservation methods (FR: deep-frozen, RL: nucleic acid preservative solution and ET: ethanol). The blue box is detailed in Suppl. material 1: Data 1. See Material and Methods for detailed explanations.

Table 1. Description of the sampling sites: site code, location, geographic references according to WGS84 system, site characteristic, aquatic ecosystem and biotic compartment are indicated.

Site Code	Location	GPS Coordinates (Latitude, Longitude)	Trophic state	Aquatic ecosystem	Biotic compartment
LC	Lim bay - Croatia	45,132529, 13,66059	mesotrophic	marine	phytoplankton
ES	Ebro bay - Spain	40,816710, 0,73077	mesotrophic	marine	phytoplankton
OF	Edian river - France	46,255750, 6,72342	oligotrophic	freshwater	phytobenthos
MS	Ebro river - Spain	40,815005, 0,51997	mesotrophic	freshwater	phytobenthos
EG	Teltow channel - Germany	52,437615, 13,32039	eutrophic	freshwater	phytobenthos
HF	Kalimenoja river - Finland	65,169722, 25,86889	humic	freshwater	phytobenthos

Freshwater phytobenthos was sampled from biofilms following the European standard (CEN 2014a) at each of the four river sites (OF, MS, EG, HF). The resulting biofilm suspension was transferred to a sterilised 1 litre bottle that was stored in a cool box with ice packs for immediate express shipment within 1 day to INRAE lab (Thonon, France) (Fig. 1). After the four river samples had been received, they were left to settle for three hours at 5 °C and concentrated by removing water supernatant to a final volume of around 900 ml of suspended biofilm per sample. Samples were then homogenised and subsampled into three sterilised bottles, one bottle of 300 ml per preservation method.

Marine phytoplankton was sampled by one vertical net haul at both marine sites (LC and ES) with a phytoplankton net (50 µm mesh size) from 15 m deep to the surface. Each net sample was suspended and evenly filtered until complete filter saturation (30 ml per filter for station LC and 60 ml per filter for station ES), on 1.2 µm cellulose (Millipore) (LC site) or GF/F glass microfibre filters (Whatman) (ES site) (Fig. 1). Net samples were well-homogenised during filtration to ensure an even distribution of the sample on each filter. For each site, 18 filters were obtained in total and were placed in marked tubes and stored in a cool box with ice packs for immediate express shipment within one day to the Center for Marine Research (CMR) lab (Rovinj, Croatia). All filters were cut in half, resulting in 12 half-filter subsamples per site and per preservation method (see details below).

Sample preservation methods

Three preservation methods were applied to phytoplankton and phytobenthos samples (Table 2, Fig. 1): 1) deep-frozen (hereafter "FR") (Visco et al. 2015), 2) preservation with a nucleic home-made acid preservative solution, based on RNAlater storage solution (Merck, Kenilworth, USA) (hereafter "RL", standing for "RNAlater" style) (Kelly et al. 2018) and 3) preservation with ethanol (hereafter "ET") (Vasselon et al. 2017a).

***FR preservation method*:** For freshwater samples, twelve 2 ml subsamples of the biofilm suspension were obtained from one 300 ml bottle under agitation for each site. Subsamples were then centrifuged, supernatant was discarded and pellets were frozen and stored at -20 °C (Suppl. material 1: Data 1C and D). For marine samples, all 12 half-filters were frozen and stored in tubes at -80 °C for each site.

***RL preservation method*:** A nucleic acid preservation solution was home-made with 3.5 M ammonium sulphate, 17 mM sodium citrate and 13 mM ethylene-diamine-tetra-acetic acid (EDTA). pH was adjusted to 5.2 using 1 M H_2SO_4 and the solution was sterilised by filtration with 0.2 µm filter. For freshwater samples, one volume of the nucleic acid preservative solution was added to one volume of sampled biofilm, for one 300 ml bottle under agitation. 24 × 2 ml subsamples of the preserved biofilm suspension were then stored for each site (Suppl. material 1: Data 1C and D). For marine samples, 2 ml of the preservative solution was added in the 12 tubes with half-filters for each site. All samples were then stored at -20 °C.

***ET preservation method*:** For freshwater samples, three volumes of 96% ethanol were added to one volume of biofilm, in order to obtain a final ethanol concentration of 70%. This was applied to one 300 ml bottle under agitation. Six 17 ml subsamples of the preserved biofilm suspension were then stored for each site (Suppl. material

Table 2. Description of the three preservation methods: method code, biotic compartment, storage conditions and material used for extraction are indicated.

Preservation method name	Preservation method code	Biotic compartment	Fixative solution	Storage temperature	Material used for preservation	Material used for extraction
Cryopreservation	FR	Phytobenthos	no	-20 °C	Pellet	Pellet
		Phytoplankton	no	-80 °C	Filter	Filter
DNA stabilization solution preservation	RL	Phytobenthos	home-made nucleic acid preservative	-20 °C	Suspended biofilm with fixative solution	Pellet
		Phytoplankton	home-made nucleic acid preservative	-20 °C	Filter with fixative solution	Filter
Ethanol preservation	ET	Phytobenthos	Ethanol (final conc. ~70%)	+4 °C	Suspended biofilm with fixative solution	Pellet
		Phytoplankton	Ethanol (final conc.~96%)	+4 °C	Filter with fixative solution	Filter

1: Data 1C and D). For marine samples, 2 ml of 96% ethanol were added in 12 tubes with half-filters. All samples were then stored in the dark at +4 °C.

In all subsampling phases for freshwater biofilm samples (Suppl. material 1: Data 1C), special attention was paid to homogenisation by: 1) permanent agitation of the solution to subsample and 2) sequential subsampling, adding solution to all subsamples in succession, 1 ml per 1 ml.

Storage duration and DNA extraction

The samples, preserved with the three methods, were further processed at six different storage durations (1 day, 1 week, 1 month, 3 months, 6 months and 1 year) during one year (Fig. 1). For each duration, two replicates were retrieved per preservation method and per site (i.e. 36 samples) and independently processed for DNA extraction.

For freshwater samples, DNA extraction was performed on biofilm pellets, either those directly preserved (FR samples, Table 2, Suppl. material 1: Data 1) or those obtained by centrifugation just before DNA extraction (RL and ET samples, Table 2, Fig. 1). In order to minimise the dilution effect of RL and ET methods (respectively ½ and ¼), DNA extraction was processed on the same total amount of 2 ml biofilm pellets for each preservation method (1 pellet for FR, 2 pooled pellets for RL and 4 pooled pellets for ET). For marine samples, wet half-filters were used directly for DNA extraction.

DNA extractions were performed using a commercial kit (Macherey–Nagel NucleoSpin Soil kit, Düren Germany) with purification columns following Vautier et al. (2020). In short, biofilm pellets and phytoplankton filters were re-suspended in lysis buffer and mechanically disrupted using ceramic beads. After proteins and undissolved sample material precipitation, supernatant with dissolved DNA was first passed through inhibitor removal columns and next through NucleoSpin Soil columns for DNA binding where PCR inhibitors were removed by efficient washing. Finally, DNA was eluted in the NucleoSpin Soil elution buffer (Tris/HCl buffer) and stored at -20 °C prior to PCR amplification and sequencing (Fig. 1). DNA extractions were processed at INRAE lab (Thonon, France) for freshwater samples and at CMR lab (Rovinj, Croatia) for marine samples. All DNA extracts from marine samples were then sent with express shipment to INRAE lab for downstream analysis at the end of the 1-year period. In total, for this 1-year experiment, 216 DNA extracts were obtained (2 replicates × 6 storage durations × 3 preservation methods × 6 sites).

DNA quality and quantity

At the end of the 1-year storage period, DNA quality and quantity were assessed on all 216 DNA extracts (Fig. 1). To evaluate the DNA quality, the 260/280 nm ratio was measured by spectrophotometry with a Nanodrop ND-1000 (Nanodrop Technologies, Wilmington, Delaware). To evaluate the DNA quantity, DNA concentration (ng/µl) was measured for each of the two DNA extract replicates with the Quant-iTTM PicoGreen dsDNA assay kit (Life Technologies, Carlsbad, California) using a microplate reader (Fluoroskan AscentTM FL; Thermo Scientific, Waltham, Massachusetts) following the manufacturer's instructions. The mean concentration value of the two replicates was used in subsequent analyses.

PCR amplification and sequencing

A 312 bp fragment of the *rbc*L chloroplastic gene was amplified from DNA extracts using Takara LA Taq polymerase and an equimolar mix of the forward primers Diat_rbcL_708F_1, 708F_2, 708F_3 and the reverse primers R3_1, R3_2 (Vasselon et al. 2017a), following the protocol of Chardon et al. (2020). For each DNA sample, a single step PCR amplification was performed in triplicate in a final volume of 25 µl. After validating PCR amplification, 19 µl of each PCR product per sample were pooled together and 50 µl of this pool were transferred into an individual well of a 96-well microplate. The resulting three plates with 216 samples were sent to "GenoToul Genomics and Transcriptomics" (GeT-PlaGe, Auzeville, France) where library preparation, final library pool and the sequencing with Illumina MiSeq System with paired-end sequencing kit (V2, 250 bp × 2) were performed (Fig. 1).

Bioinformatics

Demultiplexing and a quality check (FastQC, Andrews 2010) was performed by the sequencing service and two fastq files (forward and reverse) were provided for each sample (2 × 216 fastq files in total). Fastq files were processed using Mothur version 1.43.0 software (Schloss et al. 2009). Contigs were made by merging forward and reverse reads (make.contigs) trimmed to only the overlapping section (trimoverlap = T). Reads in every fastq file were then quality filtered (screen.seqs) by excluding sequences with an overlap shorter than 180 bp (minoverlap = 180) and all sequences with ambiguities (maxambig = 0) and mismatches (mismatches = 0). From the resulting good quality reads, exact duplicates were removed (dereplication) (unique.seqs). Subsequent bioinformatics steps on good quality reads included alignment (align.seqs) of these sequences to the reference alignment (Diat.barcode v.7 reference sequence library) and preclustering (pre.cluster) that enabled de-noising of sequences (one difference for every 100 bp of sequence was allowed). Chimera removal (chimera.vsearch) was done using the VSEARCH algorithm with default parameters (de novo chimera detection). Sequence classification (classify.seqs) was made using the naïve Bayesian Method (Wang et al. 2007) with bootstrap confidence score set to 85% and Diat.barcode v.7 as the reference sequence library (Rimet et al. 2016, 2019). Sequences classified to taxa other than diatoms (Bacillariophyta) were removed from further processing (remove.lineage). Operational Taxonomic Units (OTUs) were defined by calculating distances between

sequences (dist.seqs) and clustering (cluster) these distances, based on 0.05 difference cut-off (95% similarity) implementing the furthest neighbour algorithm. OTUs containing one single sequence (singletons) were removed. Significant correlation between replicates was detected for the number of reads (Spearman $R = 0.71$, $p < 0.001$) and the number of OTUs ($R = 0.93$, $p < 0.001$). Sample replicates were merged (merge.groups) and replicate reads were summed to retain read abundances. Using the classify.otu command, consensus taxonomy for each OTU was assigned with an 85% confidence threshold using previously classified reads taxonomy. A list of classified OTUs and their read abundances in each sample was produced. OTUs with identical highest taxonomy level were merged (merge.otus) to obtain the list of different taxa detected in the dataset. Random subsampling was performed to normalise the data with size (number of reads) set to the size of the smallest sample (14,190 reads).

Statistical analyses

Statistical analyses, as well as graphical presentations of the results, were performed using the R software version 3.6.0 (R Core Team 2017). Sequencing data import and manipulations in R were conducted with the phyloseq package. Graphical presentations were produced using ggplot2. T-tests and ANOVA were used to test for differences in DNA concentration and diatom assemblage diversity among different storage durations and preservation methods. Spearman's rho statistic was used to estimate a rank-based measure of association in correlation analyses. Patterns of sample dissimilarity were visualised using unconstrained ordinations of non-metric multidimensional scaling (NMDS). Ordinations were based on taxa presence/absence and abundance using Jaccard and Bray-Curtis indices. Further tests of the differences in the diatom assemblages used permutational multivariate ANOVA (PERMANOVA) on the normalised dataset with the adonis function in vegan version 2.4.2 (Oksanen et al. 2018) with 9999 permutations. PERMDIST with the betadisper function, also in vegan, was used to test multivariate homogeneity of group dispersions (variances). Pairwise differences within groups were determined by pairwise Adonis function (Martinez Arbizu 2020) in vegan. Differences in community composition were visualised using Venn diagrams (vegan) and taxa lists for each sample were produced with merge.otus. Finally, the presence/absence of dominant and rare (< 100 reads) taxa for different storage durations and preservation methods were analysed.

Diatom indices calculation

For freshwater river sites, we assessed their ecological quality using the Specific Pollution-sensitivity Index (SPI) (Cemagref 1982), based on species inventories (species composition and relative abundances, based on read numbers; or genus, if species level was not reached) obtained by metabarcoding. SPI values were calculated using the OMNIDIA 5 software (Lecointe et al. 1993).

Data availability

Fastq files are available at https://doi.org/10.57745/M3PESU.

Results

DNA quality and quantity

Spectrophotometry measurements confirmed good DNA quality with 260/280 nm ratios between 1.8 and 2 for all samples. Measured DNA concentrations differed among samples and ranged from 1 to 160 ng/µl (Fig. 2) with marine samples having significantly lower concentrations (mean value 7.1 ng/µl, Welch's two-sample t-test, $p < 2.2e\text{-}16$) than the freshwater ones (mean value 65.8 ng/µl). The preservation method had an effect on DNA concentration of freshwater samples, but not on that of marine samples ($F_{2,33} = 0.44$, $p > 0.05$). Freshwater samples stored in ethanol (ET) had significantly lower DNA concentrations ($F_{2,69} = 28.86$, $p < 0.001$) compared to the other two types of preserved samples (FR, RL). Storage duration had no effect on DNA concentration for either the marine samples ($F_{5,30} = 0.78$, $p > 0.05$) or the freshwater samples stored in FR and RL (Fig. 2). However, for freshwater samples stored in ET, a significant decrease in DNA concentration was observed after 3 months ($F_{5,18} = 7.55$, $p < 0.001$).

Diatom assemblage diversity

High-throughput sequencing of the 216 samples resulted in a total of 7.9 million reads. Only one sample (Ebro bay-RL-1 month) could not be sequenced successfully.

After bioinformatics processing, a total of 3.9 million (49.4%) reads were conserved with an average of 36,621 reads per sample. Read clustering (95% sequence similarity threshold) resulted in an average of 357 OTUs per sample and classification (85% bootstrap confidence score threshold) identified an average of 97 taxa per sample. Rarefaction curves indicated sufficient sequencing depth for most of the samples (Suppl. material 2: Data 2).

When all preservation methods and storage durations were considered, freshwater sites were, on average, characterised with higher number of reads (41,222 reads/sample), higher OTU and taxa richness (433 OTUs/sample and 111 taxa/sample) and higher diversity index (Shannon) values compared to marine sites (27,155 reads/sample, 199 OTUs/sample and 67 taxa/sample) (Fig. 3). The average number of OTUs and taxa across all sites is very similar for all three different preservation methods. Preservation methods had no significant impact on read numbers ($F_{2,104} = 0.533, p > 0.05$), OTU ($F_{2,104} = 0.413, p > 0.05$) and taxa ($F_{2,104} = 0.436, p > 0.05$) richness and Shannon index

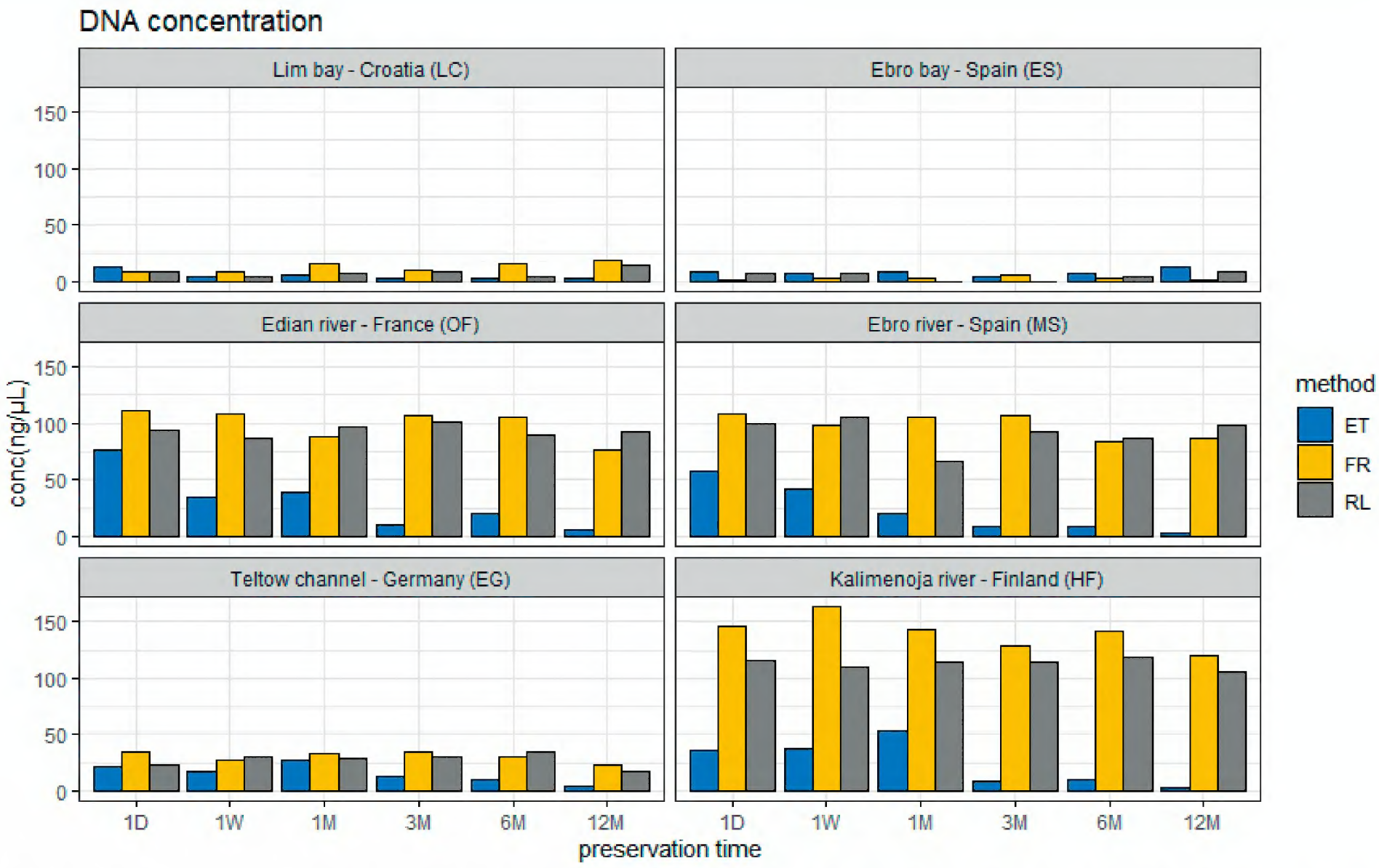

Figure 2. DNA concentrations (mean values of 2 replicates) over time (x axis: 1D – 1 day; 1W – 1 week; 1M – 1 month) for the three preservation methods (blue: ET, yellow: FR, grey: RL) for marine sites (top row) and freshwater sites (middle and last rows).

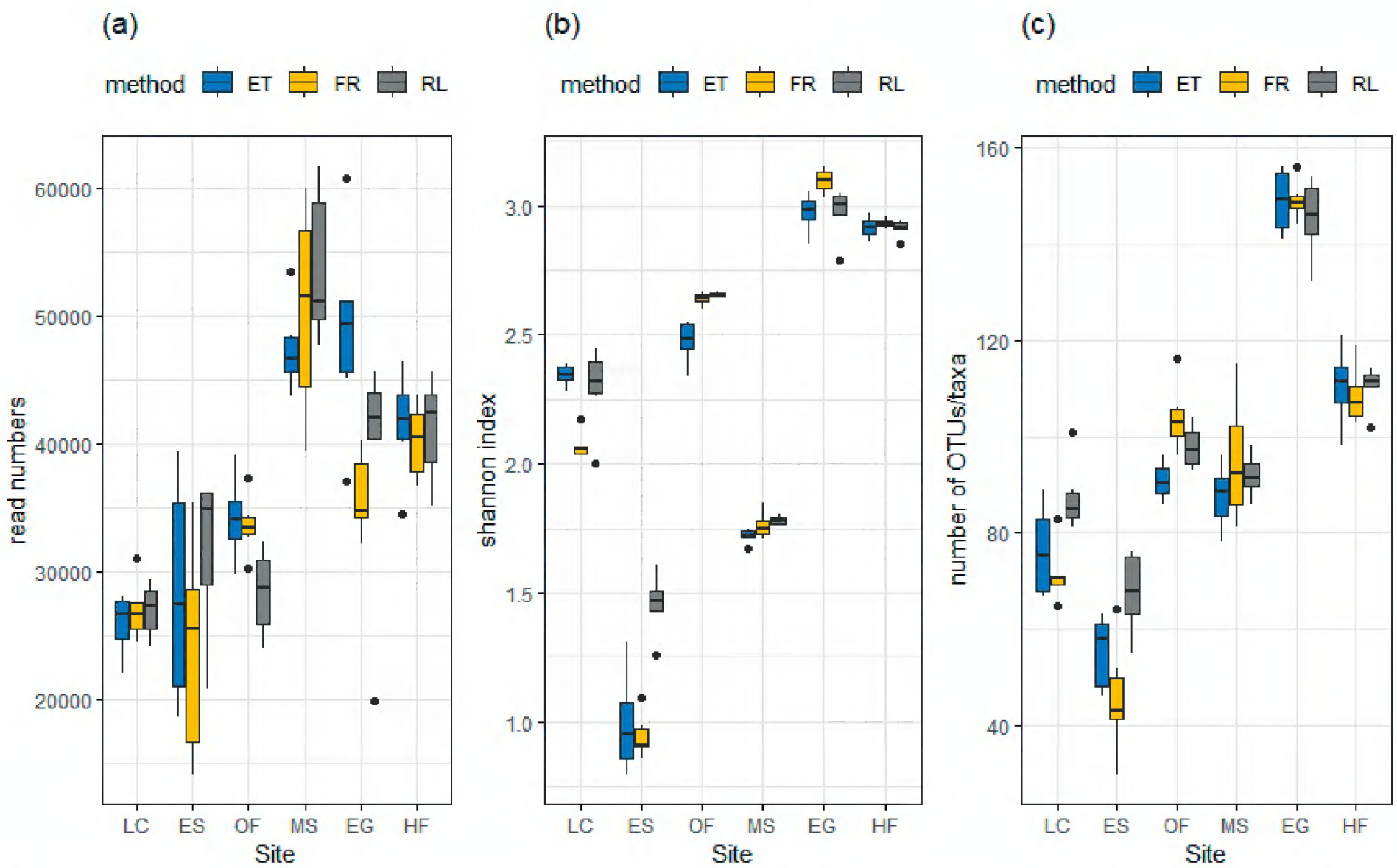

Figure 3. Diversity parameters of diatom assemblages: box plots for (**a**) read numbers, (**b**) Shannon index, (**c**) number of OTUs and (**d**) number of taxa for the three preservation methods (blue: ET, yellow: FR, grey: RL) for marine (LC, ES) and freshwater sites (OF, MS, EG, HF). Boxes represent the interquartile range, with the median indicated with a line and whiskers extending to the highest and lowest values.

values ($F_{2,104}$ = 0.439, p > 0.05) (Fig. 3). These diversity parameters did not change significantly over time (read numbers ($F_{5,101}$ = 0.38, p > 0.05), OTU ($F_{5,101}$ = 0.069, p > 0.05) and taxa ($F_{5,101}$ = 0.129, p > 0.05) richness and Shannon index values ($F_{5,101}$ = 0.025, p > 0.05)).

Diatom assemblage composition

Using the Diat.barcode reference library, 289 OTUs were assigned at species level, 77 at genus level, 21 at family level, nine at order level and two at class level. Overall, 102 different diatom genera were detected in the dataset. The diatom assemblage composition differed among sites (Suppl. material 3: Data 3), reflecting their environmental characteristics (freshwater vs. marine). The genus *Nitzschia* was the most abundant in terms of read numbers and included the highest number of species (28). About 50% of the 102 genera were represented by a single species in the dataset.

Do preservation methods and/or storage duration affect assemblage structure?

Community ordination analyses taking read abundance into account (Bray-Curtis distance) showed that the samples differed primarily according to sampling sites (PERMANOVA, pseudo$F_{5,101}$ = 573.08, R2 = 0.96, p = 0.001) (Fig. 4). Sampling sites significantly influenced the community structure when presence/absence was considered (Jaccard distance: PERMANOVA, pseudo$F_{5,101}$ = 67.45, R2 = 0.9076, p < 0.001). Since sampling sites were very different (Bray-Curtis and Jaccard distance R2 > 0.90), the effects of the preservation method and storage time on community structure were tested with site-by-site analyses. Permutation tests for homogeneity of multivariate dispersions confirmed significant differences in dispersion (p < 0.001) between sampling sites, but not between methods (p = 0.786) and time (p = 0.975).

Differences in assemblage composition between different methods can be seen in NMDS plots for each site (Fig. 4). Site-by-site analysis of read abundance (Bray-Curtis distance) and taxa presence/absence (Jaccard distance) showed that the preservation method had a significant effect (PERMANOVA, p = 0.001 for both distance matrix) on diatom assemblage structure at all sampling sites, while storage duration did not (PERMANOVA, p > 0.3 for both distance matrices). The preservation method explained on average 64% of the total variance in distance between samples, while the storage duration explained on average 11% (Table 3). Pairwise comparisons of Bray–Curtis dissimilarity between different methods (Table 4) for each site indicate significant differences in assemblages for all comparisons (p < 0.05), except FR-RL for sites MS and HF and ET-RL for site LC.

Are some taxa differentially detected?

Assemblage changes are mainly due to changes in relative abundances for abundant taxa (Suppl. material 3: Data 3) and to changes in presence-absence for low abundant taxa (not shown). There is no significant difference in the number of taxa detected between the preservation methods (ANOVA, $F_{2,104}$ = 0.436, p > 0.05) and the overall number of taxa detected for each method is around 300 per sample. In the dataset, 81% of detected taxa were shared by all the three methods, while each method is characterised with less than 5% of taxa detected only by one method (ET: 11 taxa, RL: 13 taxa, FR: 5 taxa, Fig. 5).

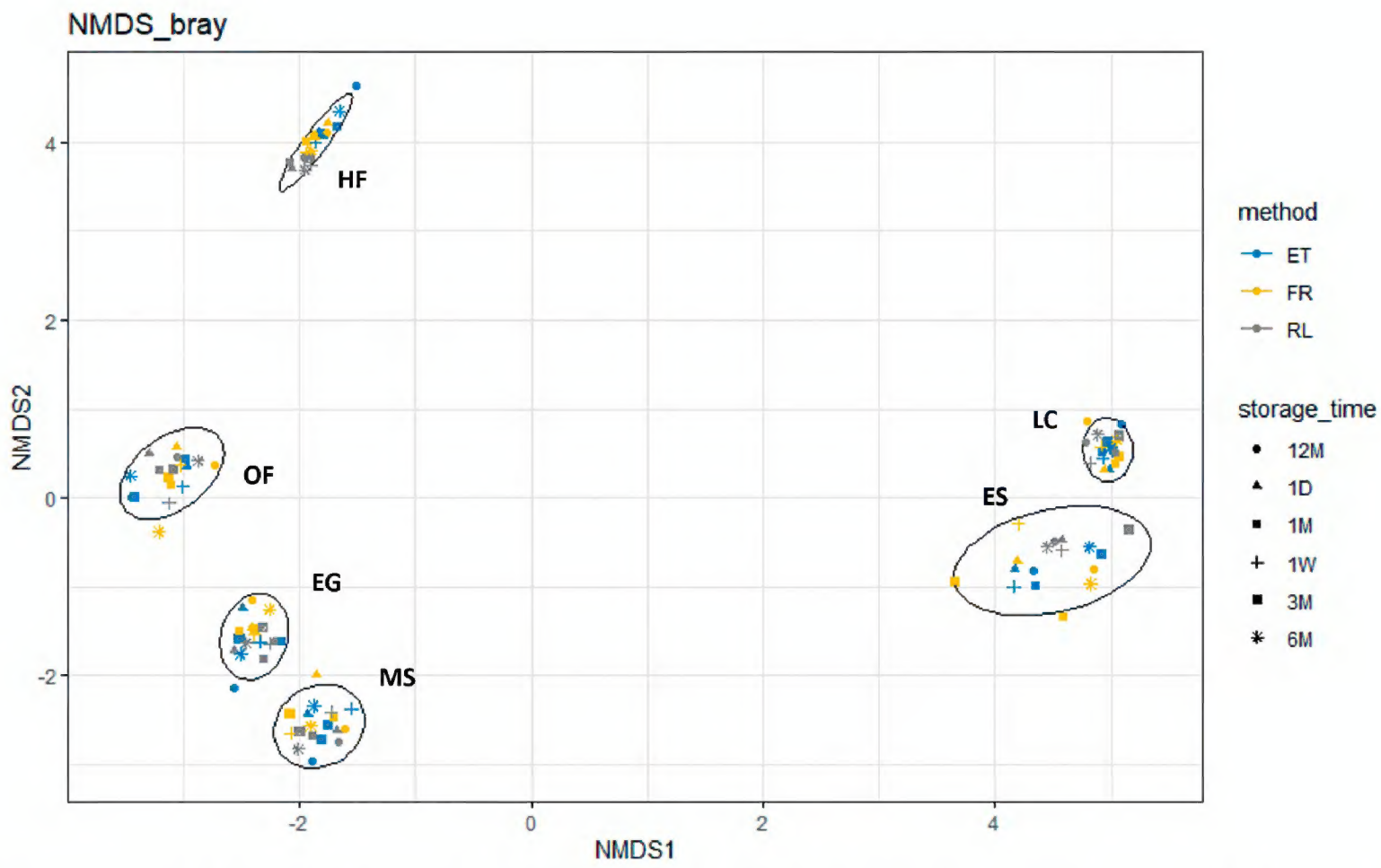

Figure 4. Non-metric multidimensional scaling (NMDS) ordinations for all sites (above) and for each site (next page), based on Bray-Curtis distances, taking read abundance into account. Samples are marked according to the preservation method (colour) and storage duration (shape). The three preservation methods (ET, FR and RL) are visualised by ellipses for the six sites (next page).

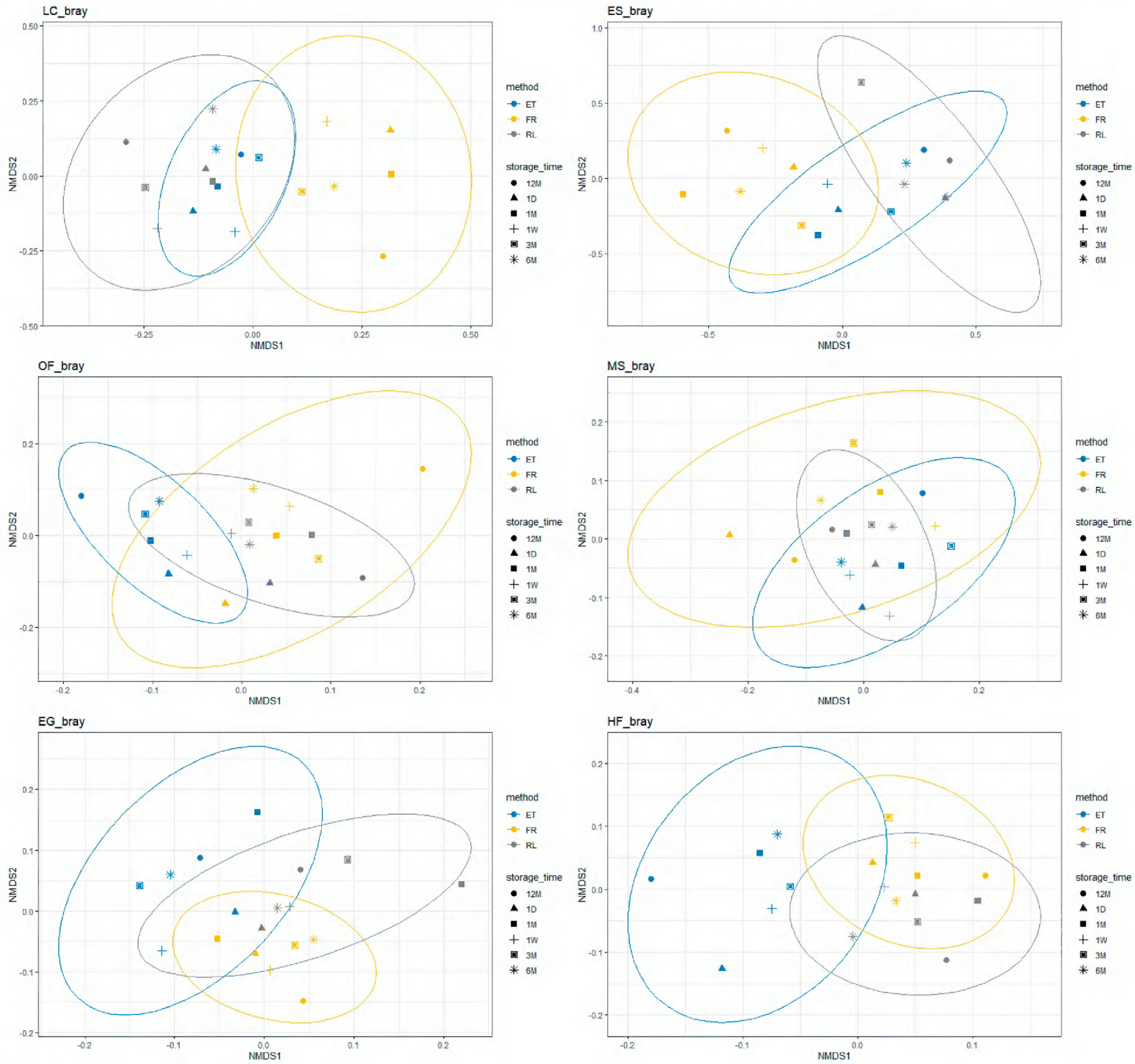

Figure 4. Continued.

Table 3. Results of PERMANOVA analysis (adonis function) of OTUs, indicating the percentage of variance (R2) explained by preservation method and storage duration and associated probability (p).

site	preservation method		storage duration	
	R2 (%)	p	R2 (%)	p
LC	0.546	0.0001	0.181	0.25
ES	0.574	0.0011	0.116	0.75
OF	0.605	0.0001	0.149	0.37
MS	0.726	0.0002	0.082	0.57
HF	0.849	0.0001	0.047	0.54
EG	0.683	0.0001	0.105	0.47

Taxa that are unique to one method are phylogenetically diverse and do not share evident ecological characteristics (cell size and shape, colony formation, habitat preference etc.). Taxa with < 100 reads were also mostly common to all three methods (36%, Fig. 5). A total of 75% of the diatom assemblage was shared in the dataset irrespective of storage durations, with those taxa unique to each duration mostly of low abundance (less than 100 reads). Only one

Table 4. Pairwise comparisons of preservation method pairs (ET-ethanol, FR-deep-frozen, RL-RNA-*later*) using pairwise.adonis function with 9999 permutations and P value adjustment method: Bonferroni.

site	pair	R2	p.adjusted
OF	ET-FR	0.576	0.015
	ET-RL	0.606	0.009
	FR-RL	0.324	0.012
MS	ET-FR	0.675	0.012
	ET-RL	0.705	0.003
	FR-RL	0.230	0.075
HF	ET-FR	0.860	0.015
	ET-RL	0.845	0.018
	FR-RL	0.173	0.096
EG	ET-FR	0.663	0.006
	ET-RL	0.415	0.018
	FR-RL	0.684	0.012
LC	ET-FR	0.592	0.012
	ET-RL	0.127	0.609
	FR-RL	0.518	0.012
ES	ET-FR	0.457	0.012
	ET-RL	0.416	0.045
	FR-RL	0.690	0.009

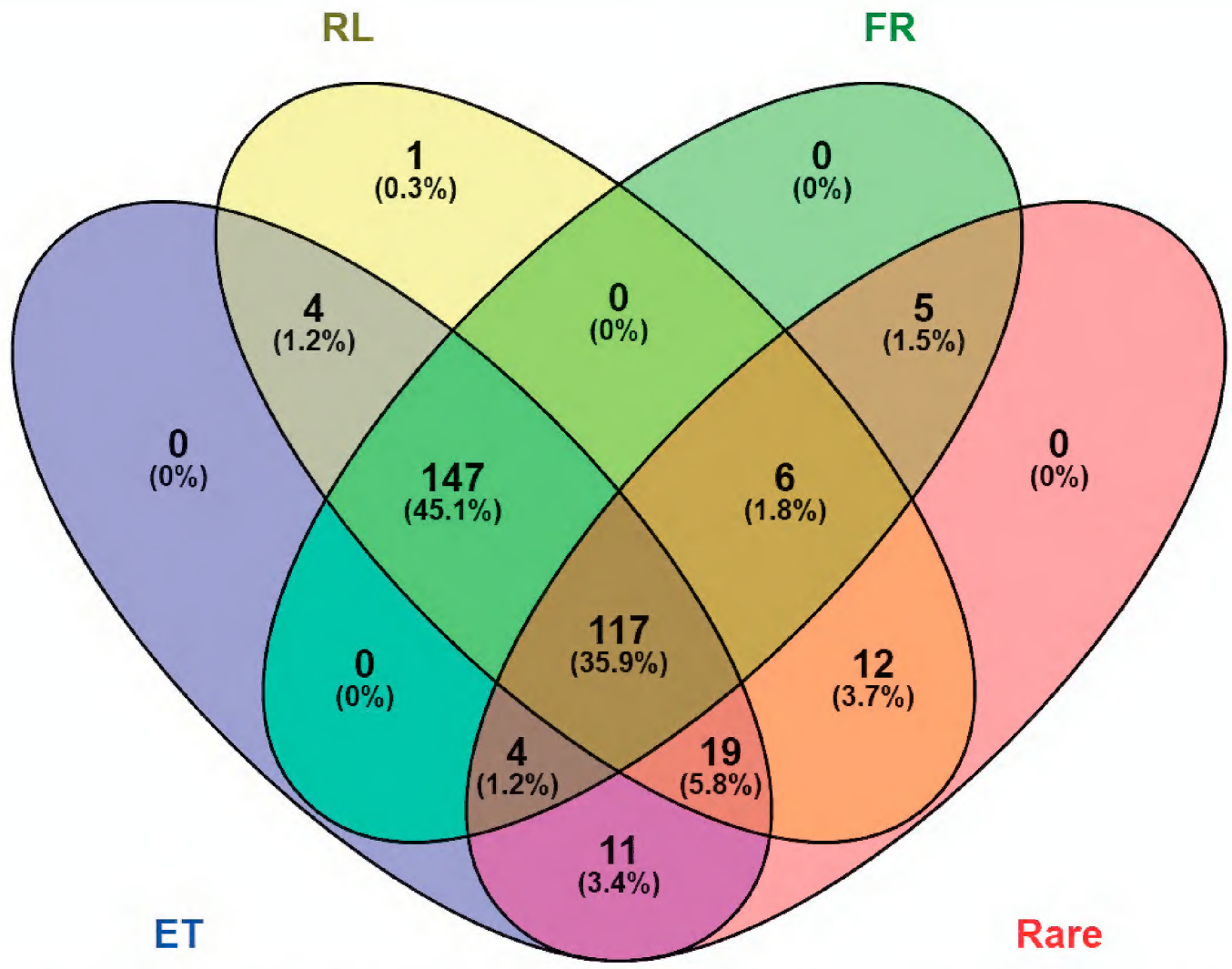

Figure 5. Comparison of the number of diatom taxa shared by the three preservation methods (ET: blue, RL: yellow, and FR: green). Rare taxa (less than 100 reads in the dataset) are presented in red.

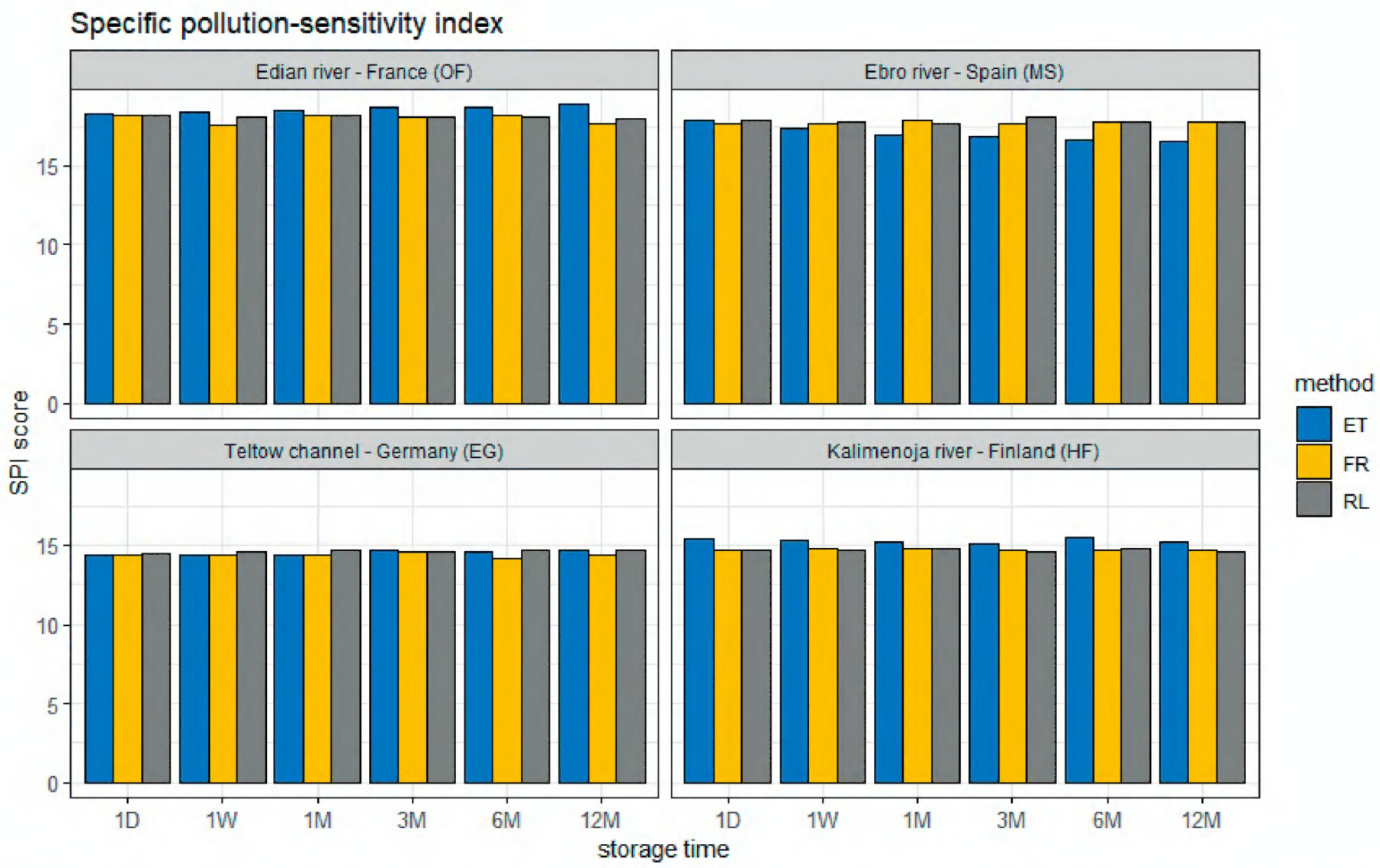

Figure 6. SPI (Specific pollution-sensitivity index) index values over the three preservation methods (ET: blue, FR: yellow, RL: grey) and the six storage durations (x axis).

species with > 100 reads was unique to the shortest duration (1 day) and the RL method: *Actinoptychus splendens*, which is a marine species detected only at the ES site (Ebro Bay, Spain). Rare taxa were often method-specific and usually appeared and disappeared over time without any obvious pattern.

Ecological quality index for freshwater sites

SPI scores were calculated for freshwater sites, based on OTUs assigned at species (73%) or genus (19%) levels and their read abundances. They ranged from 14.2–18.9 (Fig. 6). Eutrophic (EG) and humic (HF) freshwater sites

had lower mean SPI values, 14.52 and 14.90, respectively. Mesotrophic (MS) and oligotrophic (OF) sites had higher mean SPI values, 17.51 and 18.23, respectively. The influence of sampling site on SPI values was significant (ANOVA, $F_{3,68}$ = 548.8, p < 0.001). Most importantly, at each site, SPI values were very stable regardless of the preservation method (ANOVA, $F_{2,69}$ = 0.039, p > 0.05) and the storage duration (ANOVA, $F_{5,66}$ = 0.006, p > 0.05) and small variations did not translate into quality class changes.

Discussion

Identification of diatom assemblage in environmental samples through DNA metabarcoding has proved to be a reliable approach that has been successfully tested in many ecological contexts through numerous pilot studies for freshwater biomonitoring (Vasselon et al. 2019; Pérez-Burillo et al. 2020; Rivera et al. 2020; Pissaridou et al. 2021; Tapolczai et al. 2021), although less has been done for phytoplankton in marine ecosystems (Piredda et al. 2018, 2022). Sample preservation and storage are unavoidable steps in the metabarcoding workflow and this experiment has identified that all the tested preservation methods and storage durations produced reproducible quality assessments of aquatic ecosystems, although there were some differences among methods and through time for the OTUs inventories.

Sample preservation: a robust first step in diatom metabarcoding

When metabarcoding is used to assess biodiversity or ecological quality indices, based on diatom assemblages, our results show an overall robustness of the approach that is only slightly affected by the method used to preserve the samples or by the storage duration. Overall, diatom assemblage composition differed among sampling sites, rather than due to preservation methods or storage duration. Detecting an important impact of sampling site on assemblage composition is not surprising since sites were chosen to represent very diverse environments with various trophic status. Diatoms are known to have specific ecological preferences; thus, their assemblages are shaped by local environmental properties. This is the reason why these assemblages are used as proxies of phytobenthos when monitoring the ecological status of waterbodies for the WFD (Rimet 2011). In our study, little effect of preservation methods and storage duration was observed on the final end-points delivered by metabarcoding: assemblage composition, diversity indices and SPI values. Although the water chemistry was different from one site to another, this did not have an impact on preservation methods. In all cases, the site effect appears to predominate, based on their diatom assemblage structure. The longest storage duration tested here was one year after sampling, which is a valid cut-off point in the context of a monitoring program where results are awaited each year. As no significant changes have been observed throughout this period for any end-points, it suggests that longer storage may also not affect results. This was observed by the authors in previous studies where samples were stored for longer periods of time and still provided results consistent with the morphotaxonomy results for RL (Kelly et al. 2020) and for ET (Bailet et al. 2019; Kahlert et al. 2021). When evaluating future needs, managers would be mainly concerned with storage contingencies (e.g. space in freezers or cold-rooms) and need to know whether raw samples or DNA extracts have to be stored, both of which would require further experiments.

In most cases, the preservation methods we explored did not affect the quantity and quality of the DNA extracted from preserved samples. The exception is the ET method applied to freshwater samples. Preservation with ethanol seems to lead to lower DNA yield than other methods. Ethanol acts both as a killing and a preservative agent, replacing water molecules in biological tissues (Carter 2003) and has been successfully used for macroinvertebrate specimen preservation for biomonitoring (Stein et al. 2013). A minimum of 70% ethanol is necessary to ensure the fixation of the samples, preventing their degradation through time due to biotic processes. To attain that, one volume of suspended biofilms needs the addition of three volumes of 96% ethanol. For the two other methods, dilution was reduced (1/2 for RL method) or absent (FR method). However, this higher dilution in ethanol (1/4), compared to RL and FR methods, was compensated by more material (4 biofilm pellets) for DNA extraction, instead of two or one, for RL and FR, respectively. Therefore, different dilutions should not impact on the final performance of the DNA extraction. In any case, these DNA-based methods should be compared with the usual practices of morphotaxonomy which require only a small subsample of suspended biofilm, to prepare one slide on which 400 valves are identified to establish species relative abundances in the sample. Thus, a considerably smaller proportion of the sample is used for current monitoring.

Recent studies on macroinvertebrate have shown that organismal DNA is released from cells into the ethanol used for preservation during sample storage (Martins et al. 2019; Zizka et al. 2019), confirming the replacement of water molecules by ethanol in biological tissues (Carter 2003). The DNA released in ethanol has been suggested and tested as an efficient method to characterise macroinvertebrate communities (Hajibabaei et al. 2012). In our case, some DNA may have also been released from diatom cells and consequently not included in the centrifuged sample that is used for DNA extraction, thus limiting the initial DNA yield. This may have also been facilitated by the absence of a freezing step in the ET method. However, samples conserved at 4 °C will be more prone to abiotic degradations than frozen ones. Finally, bio-

films are complex samples that include many organisms (mainly bacteria, other microalgae, fungi) embedded in an exopolysaccharide matrix (Flemming and Wingender 2010) and even hosting eDNA from external organisms (Rivera et al. 2021a, 2021b). The extracted DNA is, thus, a mix of DNA from different organisms. Differential DNA release may occur during the preservation phase in ethanol, potentially adding one more dilution for diatoms.

Sample storage duration, from one day to one year, does not affect the quantity and quality of the DNA extracted from preserved samples, except when preserving freshwater samples with ethanol. In that case, a decrease of DNA concentration was observed that was marked during the three first months of preservation, but this trend did not continue over the subsequent nine months. We can hypothesise that this decrease is linked to the release of DNA from the cells to the ethanol solution. This could be further evaluated by extracting DNA from the pellets and DNA from the ethanol in parallel.

For all methods and dates, even in the "worst case" of ethanol preservation for the samples that have been stored the longest, the final end-points were not affected. Indeed, the assemblage composition is largely homogeneous at each site, whatever the method and the storage duration. The small percentage of taxa that differ from one method to the other or from one date to the other are among those that are rare (< 100 reads). When diatom metabarcoding is dedicated to the evaluation of ecological status to compare changes in assemblage structure through time and space, which is currently its main application, it is definitely an approach that is not affected by the sample conservation. If diatom metabarcoding is dedicated to the detection of rare species (e.g. invasive, endangered or toxic species), then the choice of the sample preservation method may be more critical. However, in our study, we could not identify a specific trend and derive best practice which is in line with previous observations that rare OTUs may be random and are poorly reproducible (e.g. Leray and Knowlton 2017). For that purpose, a higher sequencing depth is probably required to avoid overlooking rare species and needs to be calibrated appropriately. Moreover, in such cases, a specific study design (e.g. biological replicates, positive and negative controls) and biomolecular methods focusing on the target species may be more adapted (e.g. dPCR, qPCR).

Towards a standardised and user-friendly method

Following results from numerous pilot studies, we can be confident that diatom metabarcoding is robust and can replace or complement the current approach, based on morphotaxonomy. To do so, stakeholders call for guidelines and/or standards to accompany the deployment of the method for biomonitoring purposes (Blancher et al. 2022). General guidelines have been recently published, providing best-practices for many applications of DNA metabarcoding (Bruce et al. 2021; Pawlowski et al. 2021), representing an important milestone. However, guidelines need to be even more precise to be easily and reliably handled by the operators in a variety of contexts. Diatom metabarcoding includes several steps which require different skills (e.g. field sampling, molecular analysis, bioinformatics, ecology). For that reason, a step-by-step standardisation would be necessary to: 1) enable various actors to be involved and 2) leave the door open to future technological changes and evolutions that are numerous in the biomolecular sector.

However, methods have to remain operational and, as far as possible, user-friendly. Concerning sample conservation, depending on the context, one method or the other maybe more adapted. Methods requiring freezing or deep-freezing conservation (FR and RL) imply immediate storage and have to avoid multiple freeze-thaw cycles. To be usable, they require fast access to -80 °C or -20 °C frozen facilities. For organisations that have to organise extensive field campaigns or to access adverse environments, without access to laboratory facilities for several days, the more practicable process is the addition of a preservative solution (nucleic acid preservative or ethanol) directly in the field. In this study, we did not test the impact of the conservation time of biofilms in nucleic acid preservative, prior to centrifugation and -20 °C storage, which may be an interesting alternative. Field collections become compromised when sample processing cannot be completed within short critical time periods when essential capabilities are unavailable. However, we assume that the conservation of samples in preservative is little affected by storage temperature (frozen or room temperature) in the first week following sampling (Ladell et al. 2019), thus providing flexibility in the first sampling and storage step of the metabarcoding process. Preservation with ethanol appears to be the most practicable strategy for large or remote field campaigns, whilst storage in the dark at +4 °C facilitates the easy storage of a large number of samples. For studies orientated towards the characterisation of diatom assemblages, our results show that ethanol is well-suited to long-term preservation.

A first attempt for standardisation was done in 2018, with the publication of a technical report (CEN 2018b) in line with an existing standard developed in CEN (CEN 2014a). This technical report concerned adapting the initial sampling phase to ensure it evolved in a manner consistent with DNA metabarcoding. Indeed, the sample conservation proposed for analysis by light microscopy (e.g. Lugol's, formaldehyde) was preventing the extraction of DNA from conserved samples. Based on ongoing studies at this time, a large choice of already-tested DNA-friendly conservation methods was proposed in a technical report (CEN 2018b). From that starting point, the present study was conducted to ascertain and optimise the use of these conservation methods. Based on our results, the initial sampling step of the diatom metabarcoding process appears to be robust whatever the preservation method used and whatever the storage duration, up to one year and probably longer. Consequently, preservation method choice in future diatom metabarcoding studies dedicated

to ecological quality assessment could be primarily based on experimental design, field and lab facilities, shipping constraints and available funding and less on necessary choice of one optimal sample preservation method.

Such robustness has been already observed for other steps of diatom metabarcoding: DNA extraction methods (Vasselon et al. 2017b), PCR amplification (Vasselon et al. 2021), bioinformatic processing (Bailet et al. 2020; Rivera et al. 2020). In all these studies, final end-points, especially ecological quality indices, were seldom affected by changes in methods. An open-access reference library, Diat.barcode (Rimet et al. 2019) and related standards (CEN; Rimet et al. 2021) are also completing the tool-box.

Conclusion and perspectives

This study has shown that preservation method and storage duration have little effect on DNA metabarcoding results, especially when assessing diatom assemblage structure and ecological quality. Even the decrease in yield and quality of extracted DNA observed only for freshwater phytobenthos samples, preserved with ethanol, did not affect the final index values. Only low abundant taxa differed among methods and durations. Thus, preservation method choice may be important for low-density species. However, for biomonitoring purposes, freshwater ecological index values were not affected whatever the preservation method and storage duration considered (including ethanol preservation), well reflecting the site ecological status.

Diatom metabarcoding has shown to be robust enough to replace or complement the current approach, based on morphotaxonomy, paving the way to new possibilities for biomonitoring (Keck et al. 2017; Pawlowski et al. 2018; Trobajo et al. 2021). Diatom metabarcoding has proven to be less prone to identification errors, to provide high-throughput consistent data and allowing its application at large temporal and spatial scales. This is partly enabled for diatoms by the open-access to a curated reference database Diat.barcode (Rimet et al. 2019). Due to the high polymorphism of the *rbc*L barcode that is sequenced, the data produced provide a more detailed scale of observation, often at intraspecific level (Chonova et al. 2021) allowing a better understanding of the evolution of diatom assemblages through time and space and their biogeography. Moreover, new developments, such as artificial intelligence, may empower methods, making full use of the genetic signal (Tapolczai et al. 2019; Feio et al. 2020; Apothéloz-Perret-Gentil et al. 2021).

Thus, once accompanied by operational standards, the method will be ready to be deployed with confidence and prescribed for future regulatory monitoring. Since 2020, CEN has dedicated one of its working group (EN/TC 230/WG 28 - DNA and eDNA methods) to the development of new standards for genomic approaches applied to the biomonitoring of aquatic ecosystems. The results of this study will facilitate the emergence of a new standard, building on the initial technical report (CEN 2018b) and specify its contours. Indeed, based on these results, the standard need not be overly prescriptive. However, new inter-laboratory tests (e.g. Vasselon et al. 2021) will be required to develop the step-by-step standards for diatom metabarcoding. Only then will the method be ready to be deployed with confidence and prescribed in future regulatory monitoring.

Acknowledgements

This work was initiated and supported by the DNAqua-Net COST Action CA15219 'Developing new genetic tools for bioassessment of aquatic ecosystems in Europe' funded by the European Union. DNAqua-Net funded the lab exchanges of the two lead authors A. Baricevic and C. Chardon through Short-Term Scientific Missions in 2017 and 2019, respectively. This work largely benefited from the discussions with all participants to the "diatom workshop" held in Limassol (Cyprus) on 1–2 October 2019, organised by CUT and supported by DNAqua-Net, especially S. Derycke (Belgium), T. Elersek (Slovenia), S. Fazi (Italy), M. Kelly (UK), M. Kelly-Quinn (Ireland), Z. Ljubesic (Croatia), S. Theroux (USA), G. Varbiro (Hungary), M. Vasquez (Cyprus). INRAE funded the DNA sequencing at INRAE Genomics (GeT-PlaGe, Auzeville, France). RT acknowledges support of the CERCA Programme/Generalitat de Catalunya and help from IRTA technicians (D. Mateu, J.L. Costa and M. Rey) for sampling. JZ acknowledges support by the Federal Ministry of Education and Research [German Barcode of Life 2 Diatoms (GBOL2), grant number 01LI1501E]. MP acknowledges support of the CMR research vessel "Burin" crew and of the Croatian Science Foundation project: Life strategies of phytoplankton in the northern Adriatic (UIP-2014-09-6563). MK acknowledges support by the Swedish Agency for Marine and Water Management. CC, FR, VV and AB acknowledge support by the Office Français de la Biodiversité (OFB). All authors thank three reviewers for fruitful comments and Martyn Kelly for proofreading the English.

We acknowledge support by the OpenAccess Publication Fund of Freie Universität Berlin.

References

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data.

Apothéloz-Perret-Gentil L, Cordonier A, Straub F, Iseli J, Esling P, Pawlowski J (2017) Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Molecular Ecology Resources 17(6): 1231–1242. https://doi.org/10.1111/1755-0998.12668

Apothéloz-Perret-Gentil L, Bouchez A, Cordier T, Cordonier A, Guéguen J, Rimet F, Vasselon V, Pawlowski J (2021) Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. Molecular Ecology 30(13): 2959–2968. https://doi.org/10.1111/mec.15646

Bailet B, Bouchez A, Franc A, Frigerio J-M, Keck F, Karjalainen S-M, Rimet F, Schneider S, Kahlert M (2019) Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. Metabarcoding and Metagenomics 3: e34002. https://doi.org/10.3897/mbmg.3.34002

Bailet B, Apothéloz-Perret-Gentil L, Baričević A, Chonova T, Franc A, Frigerio J-M, Kelly M, Mora D, Pfannkuchen M, Proft S, Ramon M, Vasselon V, Zimmermann J, Kahlert M (2020) Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. The Science of the Total Environment 745: 140948. https://doi.org/10.1016/j.scitotenv.2020.140948

Blancher P, Lefrançois E, Rimet F, Vasselon V, Argillier C, Arle J, Beja P, Boets P, Boughaba J, Chauvin C, Deacon M, Duncan W, Ejdung G, Erba S, Ferrari B, Fischer H, Hänfling B, Haldin M, Hering D, Hette-Tronquart N, Hiley A, Järvinen M, Jeannot B, Kahlert M, Kelly M, Kleinteich J, Koyuncuoğlu S, Krenek S, Langhein-Winther S, Leese F, Mann D, Marcel R, Marcheggiani S, Meissner K, Mergen P, Monnier O, Narendja F, Neu D, Pinto VO, Pawlowska A, Pawlowski J, Petersen M, Poikane S, Pont D, Renevier M-S, Sandoy S, Svensson J, Trobajo R, Zagyva AT, Tziortzis I, van der Hoorn B, Vasquez MI, Walsh K, Weigand A, Bouchez A (2022) A strategy for successful integration of DNA-based methods in aquatic monitoring. Metabarcoding and Metagenomics 6: 215–226. https://doi.org/10.3897/mbmg.6.85652

Bruce K, Blackman RC, Bourlat SJ, Hellström M, Bakker J, Bista I, Bohmann K, Bouchez A, Brys R, Clark K, Elbrecht V, Fazi S, Fonseca VG, Hänfling B, Leese F, Mächler E, Mahon AR, Meissner K, Panksep K, Pawlowski J, Luis P, Yáñez S, Seymour M, Thalinger B, Valentini A, Woodcock P, Traugott M, Vasselon V, Deiner K (2021) A practical guide to DNA-based methods for biodiversity assessment. Advanced Books 1: e68634. https://doi.org/10.3897/ab.e68634

Carter JD (2003) The effects of preservation and conservation treatments on the DNA of museum invertebrate fluid preserved collections. Master The. Wales National Museum. Cardiff, 120 pp.

Cemagref (1982) Etude des méthodes biologiques quantitatives d'appréciation de la qualité des eaux. Agence de l'Eau Rhône — Méditerranée — Corse.

CEN (2014a) Water quality – Guidance standard for the routine sampling and pretreatment of benthic diatoms from rivers. EN 13946: 2014. Comité Européen de Normalisation, Geneva.

CEN (2014b) Water quality – Guidance standard for the identification, enumeration and interpretation of benthic diatom samples from running waters. EN 14407:2014. Comité Européen de Normalisation, Geneva.

CEN (2018a) CEN/TR 17244: Water quality. Technical report for the management of diatom barcodes. In: CEN/TC 230/WG23, Aquatic Macrophytes and Algae, 1–11.

CEN (2018b) CEN/TR 17245: Water quality. Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses. In: CEN/TC 230/WG23, Aquatic Macrophytes and Algae, 1–8.

Chardon C, Vasselon V, Vautier M, Lacroix S, Bouchez A, Rimet F, Domaizon I (2020) Diatom DNA library preparation for Illumina Miseq Sequencing using the rbcL marker gene. https://doi.org/10.17504/protocols.io.bd94i98w

Chonova T, Rimet F, Bouchez A, Keck F (2021) Revisiting global biogeography of freshwater diatoms: new insights from molecular data. ARPHA Conference Abstracts 4: e65129 4: e65129. https://doi.org/10.3897/aca.4.e65129

de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horak A, Jaillon O, Lima-Mendez G, Luke J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E, Boss E, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sullivan MB, Velayoudon D (2015) Eukaryotic plankton diversity in the sunlit ocean. Science 348(6237): 1261605–1261605. https://doi.org/10.1126/science.1261605

European Commission (2008) Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy. Official Journal of the European Union 2008(164): 25.

European Commission (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy. Official Journal 327: 1–72. [22.12.2000]

Feio MJ, Serra SRQ, Mortágua A, Bouchez A, Rimet F, Vasselon V, Almeida SFP (2020) A taxonomy-free approach based on machine learning to assess the quality of rivers with diatoms. The Science of the Total Environment 722: 137900. https://doi.org/10.1016/j.scitotenv.2020.137900

Flemming HC, Wingender J (2010) The biofilm matrix. Nature Reviews. Microbiology 8(9): 623–633. https://doi.org/10.1038/nrmicro2415

Grizzetti B, Lanzanova D, Liquete C, Reynaud A, Cardoso AC (2016) Assessing water ecosystem services for water resource management. Environmental Science & Policy 61: 194–203. https://doi.org/10.1016/j.envsci.2016.04.008

Hajibabaei M, Spall JL, Shokralla S, van Konynenburg S (2012) Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. BMC Ecology 12(1): e28. https://doi.org/10.1186/1472-6785-12-28

Kahlert M, Kelly M, Albert RL, Almeida SFP, Bešta T, Blanco S, Coste M, Denys L, Ector L, Fránková M, Hlúbiková D, Ivanov P, Kennedy B, Marvan P, Mertens A, Miettinen J, Picinska-Fałtynowicz J, Rosebery J, Tornés E, Vilbaste S, Vogel A (2012) Identification versus counting protocols as sources of uncertainty in diatom-based ecological status assessments. Hydrobiologia 695: 109–124. https://doi.org/10.1007/s10750-012-1115-z

Kahlert M, Bailet B, Chonova T, Karjalainen SM, Schneider SC, Tapolczai K (2021) Same same, but different: The response of diatoms to environmental gradients in Fennoscandian streams and lakes – barcodes, traits and microscope data compared. Ecological Indicators 130: 108088. https://doi.org/10.1016/j.ecolind.2021.108088

Kang W, Anslan S, Börner N, Schwarz A, Schmidt R, Künzel S, Rioual P, Echeverría-Galindo P, Vences M, Wang J, Schwalb A (2021) Diatom Metabarcoding and Microscopic Analyses from Sediment Samples at Lake Nam Co, Tibet: The Effect of Sample-Size and Bioinformatics on the Identified Communities. Ecological Indicators 121: 107070. https://doi.org/10.1016/j.ecolind.2020.107070

Keck F, Vasselon V, Tapolczai K, Rimet F, Bouchez A (2017) Freshwater biomonitoring in the Information Age. Frontiers in Ecology and the Environment 15(5): 266–274. https://doi.org/10.1002/fee.1490

Kelly M (2013) Data rich, information poor? Phytobenthos assessment and the Water Framework Directive. European Journal of Phycology 48(4): 437–450. https://doi.org/10.1080/09670262.2013.852694

Kelly M, Boonham N, Juggins S, Kille P, Mann D, Pass D, Sapp M, Sato S, Glover R (2018) A DNA based metabarcoding approach to assess diatomcommunities in rivers: project summary SC140024. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684491/A_DNA_based_metabarcoding_approach_to_assess_diatom_communities_in_rivers_-_summary.pdf

Kelly MG, Juggins S, Mann DG, Sato S, Glover R, Boonham N, Sapp M, Lewis E, Hany U, Kille P, Jones T, Walsh K (2020) Development of a novel metric for evaluating diatom assemblages in rivers using DNA metabarcoding. Ecological Indicators 118: 106725. https://doi.org/10.1016/j.ecolind.2020.106725

Kermarrec L, Franc A, Rimet F, Chaumeil P, Humbert JF, Bouchez A (2013) Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: A test for freshwater diatoms. Molecular Ecology Resources 13(4): 607–619. https://doi.org/10.1111/1755-0998.12105

Kermarrec L, Franc A, Rimet F, Chaumeil P, Frigerio JM, Humbert JF, Bouchez A (2014) A next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshwater Science 33(1): 349–363. https://doi.org/10.1086/675079

Ladell BA, Walleser LR, McCalla SG, Erickson RA, Amberg JJ (2019) Ethanol and sodium acetate as a preservation method to delay degradation of environmental DNA. Conservation Genetics Resources 11(1): 83–88. https://doi.org/10.1007/s12686-017-0955-2

Lecointe C, Coste M, Prygiel J (1993) "Omnidia": software for taxonomy, calculation of diatom indices and inventories management. Hydrobiologia 269: 509–513. https://doi.org/10.1007/BF00028048

Leray M, Knowlton N (2017) Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. PeerJ 5: e3006. https://doi.org/10.7717/peerj.3006

Mann DG, Vanormelingen P (2013) An inordinate fondness? The number, distributions, and origins of diatom species. The Journal of Eukaryotic Microbiology 60(4): 414–420. https://doi.org/10.1111/jeu.12047

Martinez Arbizu PR (2020) pairwiseAdonis: pairwise multilevel comparison using adonis. Package Version 0.4. https://github.com/pmartinezarbizu/pairwiseAdonis

Martins FMS, Galhardo M, Filipe AF, Teixeira A, Pinheiro P, Paupério J, Alves PC, Beja P (2019) Have the cake and eat it: Optimizing nondestructive DNA metabarcoding of macroinvertebrate samples for freshwater biomonitoring. Molecular Ecology Resources 19(4): 863–876. https://doi.org/10.1111/1755-0998.13012

Massana R, Gobet A, Audic S, Bass D, Bittner L, Boutte C, Chambouvet A, Christen R, Claverie JM, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Forn I, Forster D, Guillou L, Jaillon O, Kooistra WHCF, Logares R, Mahé F, Not F, Ogata H, Pawlowski J, Pernice MC, Probert I, Romac S, Richards T, Santini S, Shalchian-Tabrizi K, Siano R, Simon N, Stoeck T, Vaulot D, Zingone A, de Vargas C (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environmental Microbiology 17(10): 4035–4049. https://doi.org/10.1111/1462-2920.12955

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2018) vegan: Community Ecology Package. R package version 2.5–3. https://cran.r-project.org/package=vegan

Pawlowski J, Kelly-Quinn M, Altermatt F, Apothéloz-Perret-Gentil L, Beja P, Boggero A, Borja A, Bouchez A, Cordier T, Domaizon I, Leese F, Kahlert M (2018) The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. The Science of the Total Environment 637–638: 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002

Pawlowski J, Bonin A, Boyer F, Cordier T, Taberlet P (2021) Environmental DNA for biomonitoring. Molecular Ecology 30(13): 2931–2936. https://doi.org/10.1111/mec.16023

Pérez-Burillo J, Trobajo R, Vasselon V, Rimet F, Bouchez A, Mann DG (2020) Evaluation and sensitivity analysis of diatom DNA metabarcoding for WFD bioassessment of Mediterranean rivers. The Science of the Total Environment 727: 138445. https://doi.org/10.1016/j.scitotenv.2020.138445

Pérez-Burillo J, Trobajo R, Leira M, Keck F, Rimet F, Sigró J, Mann DG (2021) DNA metabarcoding reveals differences in distribution patterns and ecological preferences among genetic variants within some key freshwater diatom species. The Science of the Total Environment 798: 149029. https://doi.org/10.1016/j.scitotenv.2021.149029

Pérez-Burillo J, Valoti G, Witkowski A, Prado P, Mann DG, Trobajo R (2022) Assessment of marine benthic diatom communities: Insights from a combined morphological–metabarcoding approach in Mediterranean shallow coastal waters. Marine Pollution Bulletin 174: 113183. https://doi.org/10.1016/j.marpolbul.2021.113183

Piredda R, Claverie JM, Decelle J, de Vargas C, Dunthorn M, Edvardsen B, Eikrem W, Forster D, Kooistra WHCF, Logares R, Massana R, Montresor M, Not F, Ogata H, Pawlowski J, Romac S, Sarno D, Stoeck T, Zingone A (2018) Diatom diversity through HTS-metabarcoding in coastal European seas. Scientific Reports 8: 1–12. https://doi.org/10.1038/s41598-018-36345-9

Piredda R, Sarno D, De Luca D, Kooistra WHCF (2022) Biogeography of six species in the planktonic diatom genus *Bacteriastrum* (Bacillariophyta). European Journal of Phycology 1–12. https://doi.org/10.1080/09670262.2021.2021591

Pissaridou P, Vasselon V, Christou A, Chonova T, Papatheodoulou A, Drakou K, Tziortzis I, Dörflinger G, Rimet F, Bouchez A, Bouchez A, Vasquez MI (2021) Cyprus' diatom diversity and the association of environmental and anthropogenic influences for ecological assessment of rivers using DNA metabarcoding. Chemosphere 272: 129814. https://doi.org/10.1016/j.chemosphere.2021.129814

R Core Team (2017) A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/

Rimet F (2011) Recent views on river pollution and diatoms. Hydrobiologia 683: 1–24. https://doi.org/10.1007/s10750-011-0949-0

Rimet F, Vasselon V, A-Keszte B, Bouchez A (2018) Do we similarly assess diversity with microscopy and high-throughput sequencing? Case of microalgae in lakes. Organisms, Diversity & Evolution 18: 51–62. https://doi.org/10.1007/s13127-018-0359-5

Rimet F, Chaumeil P, Keck F, Kermarrec L, Vasselon V, Kahlert M, Franc A, Bouchez A (2016) R-Syst:diatom: An open-access and curated barcode database for diatoms and freshwater monitoring. Database 2016: 1–21. https://doi.org/10.1093/database/baw016

Rimet F, Gusev E, Kahlert M, Kelly MGMG, Kulikovskiy M, Maltsev Y, Mann DGDG, Pfannkuchen M, Trobajo R, Vasselon V,

Zimmermann J, Bouchez A (2019) Diat.barcode, an open-access curated barcode library for diatoms. Scientific Reports 9(1): 1–12. https://doi.org/10.1038/s41598-019-51500-6

Rimet F, Aylagas E, Borja A, Bouchez A, Canino A, Chauvin C, Chonova T, Čiampor F, Costa FO, Ferrari BJD, Zimmermann J, Ekrem T (2021) Metadata standards and practical guidelines for specimen and DNA curation when building barcode reference libraries for aquatic life. Metabarcoding and Metagenomics 5: 17–33. https://doi.org/10.3897/mbmg.5.58056

Rivera SF, Vasselon V, Bouchez A, Rimet F (2020) Diatom metabarcoding applied to large scale monitoring networks: Optimization of bioinformatics strategies using Mothur software. Ecological Indicators 109: 105775. https://doi.org/10.1016/j.ecolind.2019.105775

Rivera S, Vasselon V, Rimet F, Bouchez A (2021a) Aquatic biofilms can act as natural environmental DNA samplers. ARPHA Conference Abstracts 4: 64812. https://doi.org/10.3897/aca.4.e64812

Rivera SF, Rimet F, Vasselon V, Vautier M, Domaizon I, Bouchez A (2021b) Fish eDNA metabarcoding from aquatic biofilm samples: Methodological aspects. Molecular Ecology Resources 22(4): 1440–1453. https://doi.org/10.1111/1755-0998.13568

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75(23): 7537–7541. https://doi.org/10.1128/AEM.01541-09

Stein ED, White BP, Mazor RD, Miller PE, Pilgrim EM (2013) Evaluating Ethanol-based Sample Preservation to Facilitate Use of DNA Barcoding in Routine Freshwater Biomonitoring Programs Using Benthic Macroinvertebrates. PLoS ONE 8(1): e51273. https://doi.org/10.1371/journal.pone.0051273

Tapolczai K, Keck F, Bouchez A, Rimet F, Kahlert M, Vasselon V (2019) Diatom DNA Metabarcoding for Biomonitoring: Strategies to Avoid Major Taxonomical and Bioinformatical Biases Limiting Molecular Indices Capacities. Frontiers in Ecology and Evolution 7: e409. https://doi.org/10.3389/fevo.2019.00409

Tapolczai K, Selmeczy GB, Szabó B, B-Béres V, Keck F, Bouchez A, Rimet F, Padisák J (2021) The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding. Ecological Indicators 122: 107322. https://doi.org/10.1016/j.ecolind.2020.107322

Trobajo R, Burillo JP, Vasselon V, Rimet F, Bouchez A, Mann DG (2021) Species sensitivity analysis as a tool for interpreting diatom metabarcoding for WFD bioassessment. ARPHA Conference Abstracts 4: e64959. https://doi.org/10.3897/aca.4.e64959

Vasselon V, Rimet F, Tapolczai K, Bouchez A (2017a) Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). Ecological Indicators 82: 1–12. https://doi.org/10.1016/j.ecolind.2017.06.024

Vasselon V, Domaizon I, Rimet F, Kahlert M, Bouchez A (2017b) Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshwater Science 36(1): 162–177. https://doi.org/10.1086/690649

Vasselon V, Bouchez A, Rimet F, Jacquet S, Trobajo R, Corniquel M, Tapolczai K, Domaizon I (2018) Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. Methods in Ecology and Evolution 9(4): 1060–1069. https://doi.org/10.1111/2041-210X.12960

Vasselon V, Rimet F, Domaizon I, Monnier O, Reyjol Y, Bouchez A (2019) Assessing pollution of aquatic environments with diatoms' DNA metabarcoding: Experience and developments from France Water Framework Directive networks. Metabarcoding and Metagenomics 3: e39646. https://doi.org/10.3897/mbmg.3.39646

Vasselon V, Ács É, Almeida S, Andree K, Apothéloz-Perret-Gentil L, Bailet B, Baricevic A, Beentjes K, Bettig J, Bouchez A, Capelli C, Chardon C, Duleba M, Elersek T, Genthon C, Hurtz M, Jacas L, Kahlert M, Kelly M, Lewis M, Macher JN, Mauri F, Moletta-Denat M, Mortágua A, Pawlowski J, Burillo JP, Pfannkuchen M, Pilgrim E, Pissaridou P, Porter J, Rimet F, Stanic K, Tapolczai K, Theroux S, Trobajo R, Van Der Hoorn B, Ines M, Hadjilyra V, Walsh K, Wanless D, Warren J, Zimmermann J, Zupančič M (2021) The Fellowship of the Ring Test: DNAqua-Net WG2 initiative to compare diatom metabarcoding protocols used in routine freshwater biomonitoring for standardisation. ARPHA Conference Abstracts 4: e65142. https://doi.org/10.3897/aca.4.e65142

Vautier M, Vasselon V, Chardon C, Rimet F, Bouchez A, Domaizon I (2020) DNA extraction from environmental biofilm using the NucleoSpin Soil kit (Macherey-Nagel). https://doi.org/10.17504/protocols.io.bd52i88e

Visco JA, Apothéloz-Perret-Gentil L, Cordonier A, Esling P, Pillet L, Pawlowski J (2015) Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. Environmental Science & Technology 49(13): 7597–7605. https://doi.org/10.1021/es506158m

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73(16): 5261–5267. https://doi.org/10.1128/AEM.00062-07

Zimmermann J, Mora D, Tapolczai K, Proft S, Chonova T, Rimet F, Bouchez A, Fidlerová D, Makovinská J, Weigand A (2021) Metabarcoding of phytobenthos samples. In: Liška I, Wagner F, Sengl M, Deutsch K, Slobodník J, Paunović M (Eds) Joint Danube Survey 4. Scientific Report: A shared analysis of the Danube river. ICPDR – International Commission for the Protection of the Danube River, Vienna, 145–156.

Zizka VMA, Leese F, Peinert B, Geiger MF (2019) DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. Genome 62(3): 122–136. https://doi.org/10.1139/gen-2018-0048

Supplementary material 1
Data 1

Authors: Ana Baricevic, Cécile Chardon, Maria Kahlert, Satu Maaria Karjalainen, Daniela Maric Pfannkuchen, Martin Pfannkuchen, Frédéric Rimet, Mirta Smodlaka Tankovic, Rosa Trobajo, Valentin Vasselon, Jonas Zimmermann, Agnès Bouchez

Data type: png file

Explanation note: Detailed workflow of the study for phytobenthos samples. See Material and Methods for detailed explanations.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.6.85844.suppl1

Supplementary material 2
Data 2

Authors: Ana Baricevic, Cécile Chardon, Maria Kahlert, Satu Maaria Karjalainen, Daniela Maric Pfannkuchen, Martin Pfannkuchen, Frédéric Rimet, Mirta Smodlaka Tankovic, Rosa Trobajo, Valentin Vasselon, Jonas Zimmermann, Agnès Bouchez

Data type: png file

Explanation note: ASVs rarefaction curves for all 216 samples.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.6.85844.suppl2

Supplementary material 3
Data 3

Authors: Ana Baricevic, Cécile Chardon, Maria Kahlert, Satu Maaria Karjalainen, Daniela Maric Pfannkuchen, Martin Pfannkuchen, Frédéric Rimet, Mirta Smodlaka Tankovic, Rosa Trobajo, Valentin Vasselon, Jonas Zimmermann, Agnès Bouchez

Data type: tif file

Explanation note: Diatom assemblage compositions for all preservation methods and durations at each of the 6 sites.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.6.85844.suppl3